



CLUSTERING MULTI ATTRIBUTE SIMILARITY INDEX FOR CATEGORICAL DATA STREAMS

Deena Babu Mandru

Research Scholar (PhD), Department of Computer Science
Krishna University, Machilipatnam, A.P-521001.

Y.K. Sundara Krishna

Principal, Krishna University,
Machilipatnam, A.P-521001

Abstract: Data mining is an aggressively concept in information retrieval based on different attributes from different data sources. For effective data collection from data sources with respect to relevant data, one-class learning is required to perform labeled based classification with individual training sequences on attributes. In clustering, uncertain data with different data set visualization. Uncertain One Class Clustering (UOCC) with support vector machine to explore data summarization in terms of user preference. UOCC process single attributes from reliable data streams for inconsistent data. So that in this paper we propose Clustering with Multi-Attribute Framework (CMAF) to group multiple attributes to explore uncertain data from reliable data. CMAF construct matrix with different reliable attributes based on relevant features. Proposed approach defines effective data summarization for relevant data with attributes partitioning and constructs user profile based on relative attributes. Experimental results come out for proposed approach gives better and expressive results with comparison of state of art methods.

Keywords: Clustering, Uncertain One Class Classifier, Multi-attribute, Support Vector Machine, Feature Representation.

1. INTRODUCTION

Data mining is an aggressively concept in information retrieval based on different attributes from different data sources. For effective data collection from data sources with respect to relevant data one-class learning is required to perform labeled based classification with individual training sequences on attributes. For some real world data outsourcing real time data set portioning with abnormal behavioral class label instances with expensive impossible data presentation. To learn these types of collective sequences in real time data set proceedings to classify target data into distinct classifier data procedures. For variety of different applications anomaly detection, document classification image annotation and content specification for different data formations.

By observing above dialog, we discover the issue of single property on ambiguous unobtrusive components sources and thought outline considering of the customer from record purposes of intrigue sources [1][2]. Customarily endorse a structure, known as obscure one-class contemplating and thought outline thinking about structure (UOLCS) on cloudy purposes of intrigue sources, which oversees unobtrusive components of vulnerability and the idea rundown looking at in dim one-class unpretentious components sources. UOLCS incorporates two segments. First, gather an Uncertain One-Class Classifier (UOCC) to the autonomous purposes of enthusiasm into the single attribute SVM contemplating stage to make a prevalent classifier [3]. In the second perspective, we review customer's idea move from purposes of intrigue sources by making a support vectors (SVs) - based clustering strategy over the record segments [4-8].

To increase multi class label presentation with high dimensional data in real time data applications, a better system is required to process different attributes. So in this paper, we prescribe to create Clustering with Multi-Attribute Framework (CMAF) to characterize record joins in light of properties in indeterminate information streams with possible and ID formal parameters.

This paper introduces Clustering with Multi-Attribute Framework (CMAF) procedure to enhance irregular lattice to give extensively low level data set representation. It is connection based approach to access irrelevant data present in grouped data with different attributes based on similarity features. This examination just partners the hole between the methodology of data clustering and that of web association research. It furthermore extends the ability to amassing framework for particular data, which has not procured much thought in the creative works. Procedure of the cluster gathering approach showed up in figure 1 with relative parts in bunch social data bases.

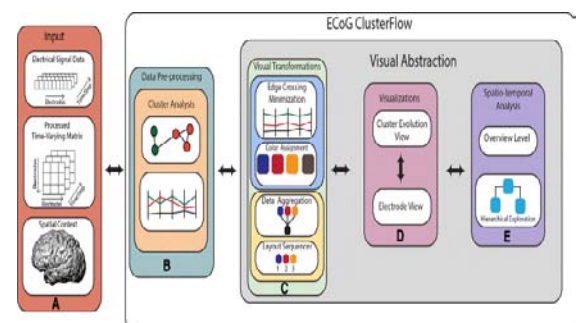


Figure 1: The cluster ensemble for feature links in uncertain cluster data.

Notwithstanding the issue of grouping specific data that is analyzed thus, the proposed structure is by and large with the end goal that it can likewise be effectively used to other data sorts.

Main objectives of proposed approach give following relational aspects

1. Briefly describe about clustering with respect to multiple attributes.
2. Respective procedure relates to clustering with different attributes.
3. Application development with respect to multi attributes partitioning.
4. Simulate the results obtained for different attributes with relations

2. BACKGROUND WORK

This section describes different author's opinion in data exploration with different attributes with different data instances from different data sources. Some of research professors and authors explore their definitions regarding data retrieval from various data sources.

Aggarwal utilizes approach based way to deal with oversee handle batch inclined and missing information. The framework centers the issue of pressing obscure articles whose reaches are outlined by likelihood thickness points of confinement and utilizations Voranoi depicts quality relations and R-Tree information to oversee inquiries in social information [9][10][11]. For efficient attribute selection and classification with Support Vector Machine to questionable data present in randomly generated for two and single attributes on center of interest points with confident relations. Like this Geo and Wang[12] process query able independent and mistake able and undefined querying data. Tang et.al defines display approaches to collect efficient querying in relational data. Based on investigation present in [13] randomly generated fixed data attributes with repeated attributes in real time attribute partition. Like in [14-18] formalize particular alliance from different random possible entity semantic relations. For efficient mining of different attributes based on procedures with real time examples. Macular et.al defines combined data from randomly generated data relations with spatial database relations.

For automatic machine learning procedure with query able for undefined information relations inspected data streams gathering through a large amount of information. For efficient attribute collection traditionally use UOCC [19][20] with machine learning to adjusted results, first collect nearest attributes based on score with different similarity in uncertain single attribute for each relation, secondly progress quadratic programming in different relations based on SVM classification with summation based refinement for each query based on single query presented for each user present in relational database[21-23]. To support different attributes with different relations in single class attributes shown in figure 2.

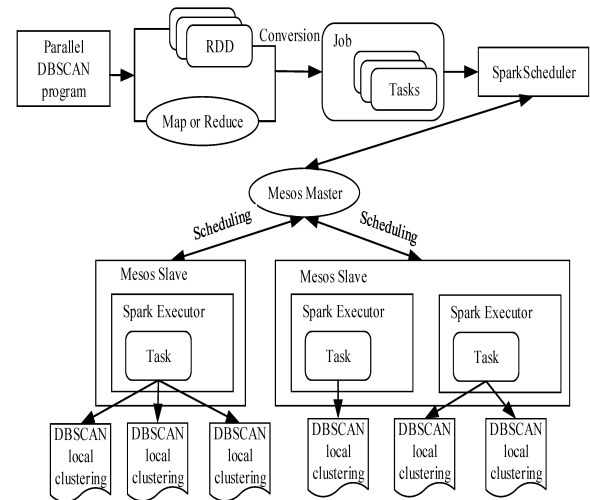


Figure 2: Concept summarization and one class learning in cluster data sets.

UOLCS structure comprises of two sections, the initial segment is to develop dubiously single node classifier from unverifiable information streams, and the remaining is idea outline machine sequence learning over the previous data streams. Two modules used in this scenario, they are 1) Single Attribute Learning 2) Topic Based Data Summarization.

2.1 Single Attribute Learning: One class learning approach [23-25] characterizes three primary modules in creating application for indeterminate information streams with plausible information streams.

For generate threshold score for instance based with local behavior using local attributes formed in different streams. In final step, for threshold based score generation to identify uncertain single attribute classification with repeated in undefined information streams. After classification different features related to information based on dimensionality in single node classification to extract relevant data streams in different data stream evaluations.

2.2. Topic Based Data Summarization: In single attribute machine learning, it is important to know different ideas and their independent relations of the client from history pieces. Traditionally, advance our bolster similarity based grouping strategy for idea synopsis gaining to different relations. Normally, for efficient data relations present in general data processing based on similarity index with different utilization, to exploit client data relations with respect to similar features in relevant data assessments[26][27]. To exploit efficient time processing with different relations for continues excess of streaming with verifiable information. Another approach uses feature based gathering strategy to consolidate thought of the customer. It initially removes features from a data knot and considers this piece as a virtual example addressed by the isolated parts, thus, the whole data streams is presented by a virtual example set, in which each virtual case addresses one data piece.

These two stages are utilized to characterize one class grouping systems for limit score figuring and characterize outline in light of order with handling examples. This strategy accomplishes one class characterization in view of occurrences as it were. So a superior framework is required for characterize with

ideal rundown traits with qualities with solid unverifiable information streams. Next section defines those relations with realistic summarization from real data sets.

3. MULTI ATTRIBUTE GROUPING METHOD

In this section, we discuss about multi-attribute clustering specification, this approach actual listening with different attributes.

3.1. Basic Procedure for Data Summarization

Let $C = (c1; c2; \dots; cN)$ be a combination of data relations with N details factors and $\gamma = (\gamma1, \gamma2, \dots, \gamman)$ Ng be a team selection with M cluster analysis, each of which is referred to as an selection individual. Each platform clustering earnings a combination of categories, $\pi_i = \{X_1^i, X_2^i, X_3^i, \dots, X_n^i\}$, such

that $\bigcup_{j=1}^{k_i} C_j^i = C$, where k_i is different selection of cluster with

different parameters. For each $x \in C$, $X(c)$ characterizes the combined brand similarity with factor c with cluster sequence. In i^{th} clustering, $X(x) = "j"$ (or " X_j^i ") if $c \in X_j^i$. This partition gives primary assets π^* of a complete set C , which contains grouped attributes with same attributes π [6][1].

So the basic cluster formation from different attribute clusters with suitable data with consensus learning functions based on results with similar attributes procedure shown in figure 3.

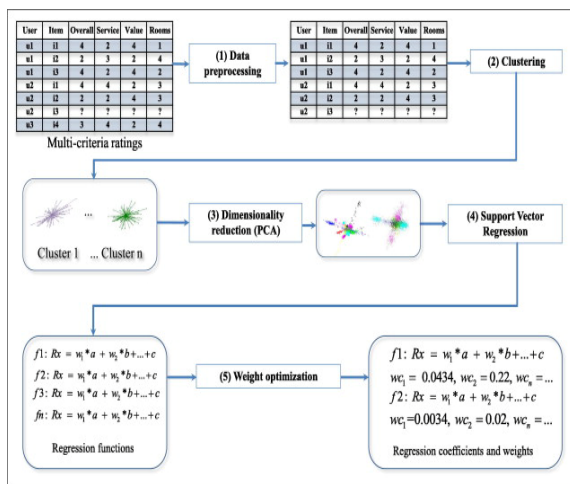


Figure 3: Cluster formation with different attributes with different cluster in similar attribute partitioning.

3.2. Grouping Creation Approach: It is the basic concept to form different attributes in combination with same relations. In clustering, individual attributes over additional data streams. Selected attributes believe many conditions with similar features based on client requirements. In this situation, selected clients operate the overall system improvement based on cluster results. Consequentially, several attributes recommended present attributes in grouping approaches with range of particular successive relational attributes. Finally successive features were used to

describe particular grouping requirements with different multi objectives.

3.3. Functions Related to Consensus: Out of overall attributes, randomly select grouped features have been designed for available information with attribute partition. Using Markov chain matrix formation has similar attributes arranged in cognitive functions. Some of the feature based approaches with cluster analysis transforms operating attributes in real time data streams for detailed categorization. In Consensus, matrix is formatted with direct and indirect labeled data.

3.4. Direct Technique: In direct approach, depending attributes are individuals for selecting relational label i.e. π and number of attributes in π_1 with multi objectives in relations for different formations using consensus function $\pi1, \pi2, \pi3, \dots, \pi m$. To provide similar attributes with grouping for random selection from different data sets. In Markova chain model matrix different formations with attributes based on Euclidian distance between all attributes in data streams [8][9].

3.5. Outlier Data Cluster for Attributes: From the procedure of direct technique with matrix formation and attribute arrangement with similar attributes in relations. Outlier formation based on attributes with multiple objectives in different consensus for grouping selected features in recent attributes to detect outlier from relations.

Algorithm for Classification Instances: For classification instances on attributes in real world synthetic data sets. Procedure of the classification instances with different attributes is as follows:

Algorithm:

Input: Records ($R=r1, r2, r3, \dots, r_m$), Queries ($P=p1, p2, p3 \dots p_n$)

Output: Possible Enhanced Time, $PT=pt1, pt2, pt3 \dots pt_n$.

Collect pattern results for submitted query (P).

1. Initial query sequence is $p_s=0$.
2. After receiving query with in time records the probability is $p(q/t)$.
3. Classify the pattern sequences ($C p_s$) based on existing ones.
4. Compute query occurrence based on patterns with p_s .
5. Query histogram with publication time (t) and with increased Records (R)
6. Compare each p_s probability time with original pattern time efficiency.
7. Save probability of p_s in published records
8. Return each probability $p(q/t)$ for each record (d) to query (p)

Table 1: Procedure of the proposed approach algorithm with relevance and label representation in training and testing usual representation.

Clustering with Multi-Attribute Framework (CMAF) procedure shown in table 1, it is step by step process for query attribute partition with multiple objects in relational data streams.

4. PERFORMANCE EVALUATION

In this section, we formulate the performance analysis of proposed approach MAMG with comparison of UOCC on real data streams. To develop this application, we use JDK and Net beans for user interface construction to upload data sets and perform single attribute classification and multi attribute object classification from real time data streams. Sample data sets shown in figure 4.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	LIVIO	Hemang	Male	#####	livio_hem208@a1700-91-829876	71	5	57	62	25	23	48		
2	TILAN	Apparajit	Male	#####	tilan_appfbc38614-91-837531	89	5	18	23	8	1	9		
3	Rajat	Vajraman	Male	6/5/2003	rajat_76814cc7ec98a-91-838282	78	39	39	78	29	1	30		
4	Madhulal	Y	Female	#####	madhulal1125a00191-91-752958	2	77	99	170	59	4	63		
5	Kanchan	CHARLY	Female	#####	charly@c91933ca23-NA	0	0	0	0	0	0	0		
6	Sai	D	Female	#####	d_236@grba0020f0c-91-975986	19	2	4	6	73	86	159		
7	Urmila	CORREN	Female	#####	urmila_co_bfa58b51-91-984556	4	9	9	18	3	2	5		
8	Shevani	Q	Female	#####	q@yahoo_677ea1u25-91-837287	98	4	17	21	8	58	66		
9	Anuna	KELECHI	Female	#####	anuna_kel33c9495d-91-795143	2	71	1	72	4	9	13		
10	RUD	MATTI	Male	#####	rud_mate_47b5e1ef-91-766273	46	2	8	10	6	8	14		
11	Sankari	ABUL	Female	#####	sankari@y7610e223-91-997797	5	57	97	154	76	83	159		
12	Jitendra	W	Male	#####	jitendra_w8294d3d-91-912164	94	78	57	135	6	6	12		
13	Trupti	V	Female	#####	trupti_v_53c1812802-91-984257	33	8	93	101	25	3	28		
14	BOSTON	R	Male	#####	boston_b1252be354-91-762993	5	86	26	112	1	6	7		
15	Sacchidan	O	Male	9/5/2006	sacchidan_9b6df1d0-91-894257	99	1	8	9	62	86	148		
16	tanaya	RYLSA	Female	1/5/2013	tanaya_ry_f941f947-NA	0	0	0	0	0	0	0		
17	SHERIFF	G	Male	#####	g_157@gr01667c7c7-91-974878	2	8	47	55	29	32	61		
18	Vasanta	DALEY	Female	2012-0-11	vasanta_d0e5f7c14-NA	0	0	0	0	0	0	0		
19	MARLOWE	JOHN JOE	Male	1/7/2013	marlowe_73d71067-91-931812	4	4	78	82	9	78	87		
20	DILLON	L	Male	#####	dillon_l013ab0d11e-91-761106	5	5	60	74	80	7	96		
21	Sushma	Q	Female	#####	sushma_q394d0429-91-943599	1	65	9	74	92	4	96		
22	CHESENEY	M	Male	3/2/2010	m@gmail_a5c80375a-91-952178	38	69	28	91	3	3	6		
23	Chandoni	Palashius	Female	3/3/2012	chandoni_4e91712f-91-753396	57	1	52	53	5	14	19		
24	Nandita	Abhay	Female	2012-4-31	nandita_a_b1091d9f-91-837153	5	22	17	39	8	32	40		
25	TAMMO	O	Male	#####	o@yahoo_el803de33-91-774711	79	11	14	25	2	3	5		

Figure 4: Sample data sets with high dimensions and attributes in data streams.

As shown in figure 4, we are taken data sets from different social networks using API (Application Programming Interface) developed by java for extracting relational data with multi attributes. And also, we take data with relational attributes shown in figure 5.

I	VE_TOTAL	PERSONS	PEDS	NO_LANE	SP_LIMIT	FATALS	DRUNK_DR
2	1	2	1	3	0	1	0
3	1	1	0	2	0	1	1
4	3	3	0	2	0	1	0
5	2	2	0	2	0	1	0
6	1	6	0	2	0	3	0
7	2	8	0	2	15	1	0
8	1	2	1	2	15	1	1
9	1	2	1	2	15	1	2
10	1	1	0	2	20	1	1
11	1	2	0	2	20	1	0
12	1	4	0	2	20	1	0
13	1	1	0	9	25	1	1
14	1	3	2	6	25	1	0
15	2	4	0	5	25	1	1
16	2	6	0	4	25	1	1
17	1	4	0	4	25	2	0
18	1	2	1	4	25	1	0
19	3	4	0	4	25	1	1
20	1	2	1	4	25	1	0
21	1	2	1	4	25	1	2
22	4	6	1	4	25	1	0
23	2	6	0	4	25	1	0
24	1	7	1	4	25	1	0
25	3	3	0	4	25	1	0

Figure 5: Accident data with different attributes.

We will take accident data from different areas in recent contribution of relational attributes as shown in figure 5.

Implementation Procedure: Implementation of MAMG with different multi attributes follows following procedure for data processing. The level of affirmation of the express data as demonstrated in beneath with component of the taking care of data focuses brings about ongoing data puts as takes after with contains spans $[V_{L_C(i,\beta)}, U_{L_C(i,\beta)}]$ for the mean as $L_{C(i,\beta)}$ with validity criteria C as follows:

$$V_{L_C(i,\beta)} = L_{X(i,\beta)} - 2.34 \frac{S_{X(i,\beta)}}{\sqrt{n}} \text{----- (1)}$$

$$U_{L_C(i,\beta)} = L_{V(i,\beta)} + 2.34 \frac{S_{V(i,\beta)}}{\sqrt{n}} \text{----- (2)}$$

As appeared in the above figure $S_{X(i,\beta)}$ is standard deviation of the validity index C cross n runs for a grouping technique I and data set β . Contrast with the preparing of prior strategies and proposed application improvement ascertained by utilizing better execution when forms clusters

$$B_{X(i)} = \sum_{\forall \beta \in DT} \sum_{\forall i^* \in CN, i^* \neq i} better_C^\beta(i, i^*), \text{----- (3)}$$

$$better_C^\beta(i, i^*) = \begin{cases} 1 & \text{if } L_{XC(i,\beta)} > U_{XC(i^*,\beta)} \\ 0 & \text{otherwise} \end{cases} \text{----- (4)}$$

Similarly, the number of times that one method $i < CM$ is significantly worse than its competitors, $WC(i)$, in accordance with the validity criterion C, can be computed as:

$$W_{C(i)} = \sum_{\forall \beta \in DT} \sum_{\forall i^* \in CM, i^* \neq i} worse_C^\beta(i, i^*), \text{----- (5)}$$

$$worse_C^\beta(i, i^*) = \begin{cases} 1 & \text{if } U_{XC(i^*,\beta)} < L_{XC(i,\beta)} \\ 0 & \text{otherwise} \end{cases} \text{----- (6)}$$

By using these equations from 1-6, gives implementation definitions for multi attributes for data processing. The efficiency of MAMG to define multi features in real time attributes.

4.1. Experimental Results: As shown in table 1, proposed approach gives accuracy of real world entity in different data sets like Accident, diabetes, Economy Ratings, Student marks performance in different formations. Sample results for after performing proposed approach for different data attributes shown in figure 6.

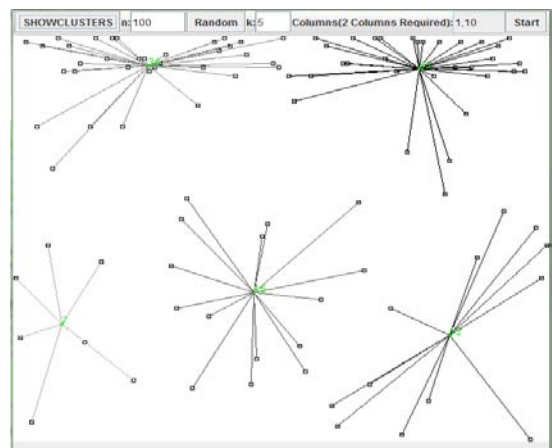


Figure 6: Multi attribute data representation with different data items.

As shown in figure 6, first upload data sets to our proposed approach for attribute selection with different relations and then process each attribute as data point, then perform Euclidian distances between attributes for efficient data classification with multi attributes. Cluster formation based on different attributes with distance calculation shown in figure 7.

```

-->Centroid (442,42) Aligning with (1085.0,18.0) @ (493,99)
-->Centroid (442,42) Aligning with (1086.0,7.0) @ (499,40)
-->Centroid (475,379) Aligning with (1087.0,52.0) @ (505,281)
-->Centroid (442,42) Aligning with (1088.0,1.0) @ (510,8)
-->Centroid (442,42) Aligning with (1089.0,9.0) @ (516,51)
-->Centroid (442,42) Aligning with (1090.0,5.0) @ (522,29)
-->Centroid (442,42) Aligning with (1091.0,2.0) @ (528,13)
-->Centroid (475,379) Aligning with (1092.0,41.0) @ (534,222)
-->Centroid (442,42) Aligning with (1093.0,0.0) @ (540,3)
-->Centroid (442,42) Aligning with (1094.0,8.0) @ (545,45)
-->Centroid (442,42) Aligning with (1095.0,6.0) @ (551,35)
-->Centroid (442,42) Aligning with (1096.0,8.0) @ (557,45)
-->Centroid (475,379) Aligning with (1097.0,95.0) @ (563,511)
-->Centroid (475,379) Aligning with (1098.0,45.0) @ (569,243)
-->Centroid (475,379) Aligning with (1099.0,57.0) @ (575,307)
-->Centroid (475,379) Aligning with (1100.0,49.0) @ (581,265)
-----
Edge Cluster Walk
[225,Cluster : 1 @ (442,42) clusters count are : 32]
[174,Cluster : 4 @ (147,38) clusters count are : 34]
[123,Cluster : 3 @ (262,325) clusters count are : 15]
[62,Cluster : 2 @ (475,379) clusters count are : 12]
[19,Cluster : 5 @ (53,365) clusters count are : 7]
Cluster Ensemble runs in 1.039 secs
    
```

Figure 7: Cluster formation for different edges for grouping multi attributes.

Time efficiency for our proposed approach shown in figure 8, different data sets like accident, diabetes, with multi attributes in recent feature selection with randomly progress real time data streams.

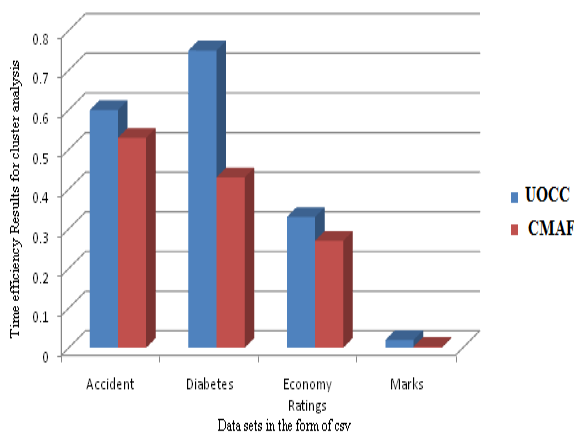


Figure 8: Experimental results comparison of both UOCC and CMAF in real time data sets.

Moreover, MAMG works continually superior to its rivals with all extraordinary determination estimations, while CO+SL give off an impression of being the slightest viable. Understand that a greater determination results in an improved flawlessness, yet with the exchange off of runtime.

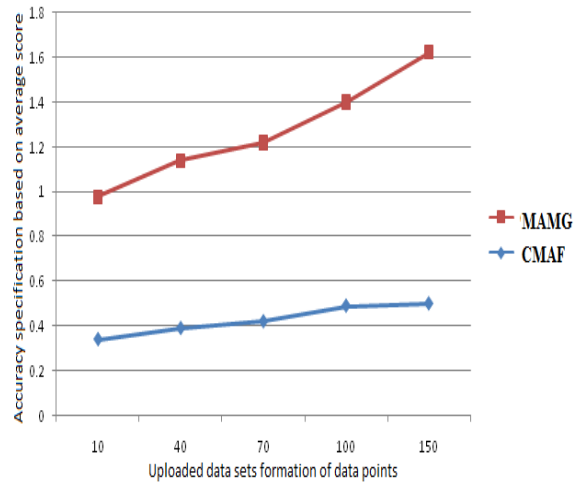


Figure 9: Accuracy measure based on generated data point processing real time datasets using two relational approaches.

Efficiency results appear with CMAF with different multi objects. To begin with, we make the restricted score to catch the territory vulnerability in light of each outline's near to data perform, and after that create a dicey one-class classifier by consolidating the vulnerability information into a CMAF with SVM-based examining framework. Second, we make upgrade vectors-based accumulation system to compress the comprehension of the purchaser over the history pieces. A wide evaluation has uncovered that our unverifiable one classification examining can get a magnificent execution and is less delicate to object regarding the standard one-class SVM. The evaluations furthermore show that the help vectors-based accumulation method can well diminish the comprehension of the customer regarding underscore focused gathering method for idea rundown learning.

5. CONCLUSION

In this paper, we propose and develop Clustering with Multi Attribute Framework for data exploration in different data sources with multi object orientation in cluster relational data bases. This paper presents novel CMAF to categorize data based on different attributes from multi dimensional data sources. It constructs and transforms matrix formation into attribute partition based on graph procedure. Our experimental results give efficient and effective approaches to configure data sets to measure attributes and combine those attributes using link based methodology It gives effective results in multi attribute combination from cluster relational data source with semantic data structure with similarity measures with feature partition. As further improvement, our proposed approach CMAF is to be implemented in different types of real time data streams using advanced machine learning algorithms. We also design best evaluation method to extract data based on multi attributes.

REFERENCES

- [1] K. Balasubramanian, P. Donmez, and G. Lebanon. Unsupervised supervised learning ii: Margin-based classification without labels. *JMLR*, 12:3119–3145, 2011.
- [2] M. Belkin, P. Niyogi, and V. Sindhwani. Manifold regularization: a geometric framework for learning from labeled and unlabeled examples. *JMLR*, 7:2399–2434, 2006.
- [3] J. Bi and T. Zhang. Support vector classification with input data uncertainty. In *NIPS 17*, 2004.
- [4] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58(3):Article 11, 2011.
- [5] Y. Chen, X. S. Zhou, and T. S. Huang. One-class svm for learning in image retrieval. In *Proc. ICIP*, 2001.
- [6] K. Crammer and G. Chechik. A needle in a haystack: Local one-class optimization. In *Proc. ICML*, 2004.
- [7] E. Elhamifar, G. Sapiro, and R. Vidal. See all by looking at a few: Sparse modeling for finding representative objects. In *Proc. CVPR*, 2012.
- [8] R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman. Learning object categories from internet image searches. *Proceedings of the IEEE*, 98(8):1453–1466, 2010.
- [9] W. Gander, G. H. Golub, and U. von Matt. A constrained eigenvalue problem. *Linear Algebra and its Applications*, 114/115:815–839, 1989.
- [10] G. Gupta and J. Ghosh. Robust one-class clustering using hybrid global and local search. In *Proc. ICML*, 2005.
- [11] J. Kim and C. D. Scott. Robust kernel density estimation. *JMLR*, 13:2529–2565, 2012.
- [12] J. Krapac, M. Allan, J. Verbeek, and F. Jurie. Improving web image search results using query-relative classifiers. In *Proc. CVPR*, 2010.
- [13] Wei Liu†, Gang Hua‡, John R. Smith, “Unsupervised One-Class Learning for Automatic Outlier Removal”, 2014 IEEE Conference on Computer Vision and Pattern Recognition.
- [14] Smith Tsang†, Ben Kao†, Kevin Y. Yip‡, Wai-Shing Ho†, Sau Dan Lee, “Decision Trees for Uncertain Data”, *IEEE Trans. Knowl. Data Eng.*, 1993.
- [15] C. L. Tsien, I. S. Kohane, and N. McIntosh, “Multiple signal integration by decision tree induction to detect artifacts in the neonatal intensive care unit,” *Artificial Intelligence in Medicine*, vol. 19, no. 3, 2000.
- [16] Bo Liu, Yanshan Xiao, Philip S. Yu, “Uncertain One-Class Learning and Concept Summarization Learning on Uncertain Data Streams”, *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, VOL. 26, NO. 2, FEBRUARY 2014.
- [17] F. Bovolo, G. Camps-Valls, and L. Bruzzone, “A Support Vector Domain Method for Change Detection in Multitemporal Images,” *Pattern Recognition Letters*, vol. 31, no. 10, pp. 1148-1154, 2010.
- [18] L. Chen and C. Wang, “Continuous Subgraph Pattern Search over Certain and Uncertain Graph Streams,” *IEEE Trans. Knowledge and Data Eng.*, vol. 22, no. 8, pp. 1093-1109, Aug. 2010.
- [19] X. Lian and L. Chen, “Similarity Join Processing on Uncertain Data Streams,” *IEEE Trans. Knowledge and Data Eng.*, vol. 23, no. 11, pp. 1718-1734, Nov. 2011.
- [20] B. Geng, L. Yang, C. Xu, and X. Hua, “Ranking Model Adaptation for Domain-Specific Search,” *IEEE Trans. Knowledge and Data Eng.*, vol. 24, no. 4, pp. 745-758, Apr. 2012.
- [21] S. Hido, Y. Tsuboi, H. Kashima, M. Sugiyama, and T. Kanamori, “Statistical Outlier Detection Using Direct Density Ratio Estimation,” *Knowledge and Information Systems*, vol. 26, no. 2, pp. 309-336, 2011.
- [22] S.V. Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*. SIAM Press, 1991.
- [23] S.R. Gunn and J. Yang, “Exploiting Uncertain Data in Support Vector Classification,” *Proc. 14th Int’l Conf. Knowledge-Based and Intelligent Information and Eng. Systems*, pp. 148-155, 2007.
- [24] B. Jiang, M. Zhang, and X. Zhang, “OSCAR: One-Class SVM for Accurate Recognition of CIS-Elements,” *Bioinformatics*, vol. 23, no. 21, pp. 2823-2828, 2007.
- [25] R. Jin, L. Liu, and C.C. Aggarwal, “Discovering Highly Reliable Subgraphs in Uncertain Graphs,” *Proc. ACM SIGKDD Int’l Conf. Knowledge Discovery and Data Mining*, pp. 992-1000, 2011.
- [26] B. Li, K. Goh, and E. Chang, “Using One-Class and Two-Class SVMs for Multiclass Image Annotation,” *IEEE Trans. Knowledge and Data Eng.*, vol. 17, no. 10, pp. 13330-1346, Oct. 2005.
- [27] B. Kao, S.D. Lee, F.K.F. Lee, D.W. Cheung, and W. Ho, “Clustering Uncertain Data Using Voronoi Diagrams and R-Tree Index,” *IEEE Trans. Knowledge and Data Eng.*, vol. 22, no. 9, pp. 1219-1233, Sept. 2010.